

JAP20 Rec'd PCT/PTO 08 MAY 2006

CALL SET-UP SYSTEMS

This invention relates to call set-up systems.

- 5 Call set-up systems are known in which call set-up involves a plurality of call agents associated with respective packet-switched networks which are connected to each other by means of NAT (network address translation) devices, also known as address translators. The NAT devices define addresses within one network which provide a connecting path to the other network to which it connects. The call set-up devices define
- 10 a series of addresses, including those of NAT devices, via which the media packets of the call are sent. In a call between two user terminals, for example, there are two paths between the two terminals, one in each direction. Typically, the media call takes place between one or more private networks joined by the internet.
- 15 The invention is particularly concerned with voice calls. There is increasing interest in using IP (Internet Protocol) for voice in place of the usual circuit-switched telecommunications network, because there are maintenance savings if one network can be used for two different functions.
- 20 One known protocol for initiating voice calls via the internet is Session Initiation Protocol (SIP-RFC 3261), although it can also be used for initiating calls using other interactive media such as video or games. This protocol is adapted for use in the case of calls involving private networks and the internet.

Figure 1 shows the set-up of a call using SIP signalling, between two networks.

A call is made from user agent P in network 1, for example, a private network, to user X (not shown) in the central network 2, for example, the internet. It is forwarded on behalf
5 of user X by call agent S to user agent V, which is in the same private network 1 as the caller.

All SIP signalling messages are shown in Figure 1, but the text of messages is not shown. SIP messages are standard messages whose format can be seen from IETF call
10 examples documents.

User agent P initiates a call by sending an SIP Invite message to its local call agent Q. This message contains a session description (Session Description Protocol - RFC 2327) indicating the media characteristics and the address (1.1.1.1) at which user agent P
15 wishes to receive media packets. For convenience, only the last two segments of this address, and of other addresses, are shown in Figure 1.

Call agent Q responds with a SIP '100: trying' message.

20 Call agent Q determines that the destination, X, of the call is in the central network 2, and that this is reached via a network NAT device R which it controls. It therefore opens a pinhole in NAT device R to permit a media flow from the central network to the media address of user agent P (1.1.1.1). The address returned to call agent Q by NAT R (2.2.2.1) is then used in an Invite message sent by call agent Q to call agent S in the

central network. (A pinhole is a path through the NAT device that is opened specifically for a single media flow. The term 'pinhole' emphasises that only packets with a particular combination of source and destination addresses will pass through the NAT device, while other packets that do not match a pinhole will be blocked.)

5

Call agent S responds with a '100: trying' message.

Call agent S determines that user X has requested that their calls be forwarded to user agent V, and that the new destination is reached via call agent U in private network 1.

10 User X could be a home telephone number accessible via the internet, and the user could have set up an arrangement for their calls to be forwarded to their office in the private network 1. Call agent S therefore passes on the Invite message with a changed URI (Uniform Resource Identifier), in this case, the name of the communications resource defining the destination of the call, to call agent U.

15

Call agent U responds with a '100: trying' message.

Call agent U recognises that the call has arrived from a different network via a NAT device T that it controls. It therefore opens a pinhole in NAT device T to permit a media

20 flow from the edge network 1 to the media address within the central network (2.2.2.1).

The address returned by NAT device T (1.1.1.3) is then used in the Invite message sent to user agent V.

The call has now reached its destination. User agent V responds with a '180: ringing' message that is passed back to the caller via the chain of call agents.

When the call is answered, user agent V sends a '200: OK' message back to call agent
5 U. The message contains a session description indicating the media characteristics and the address (1.1.1.4) at which user agent V wishes to receive media packets.

Call agent U recognises that this message is for a call that arrived via NAT device T that it controls. It therefore opens a pinhole in NAT device T to permit a media flow from
10 the central network 2 to the media address within the edge network (1.1.1.4). The address returned by NAT device T (2.2.2.2) is then used in the OK message passed on to call agent S.

The OK message is passed back by call agent S to call agent Q.

15

Call agent Q recognises that this call passes through NAT device R that it controls. It therefore opens a pinhole in NAT device R to permit a media flow from the edge network 1 to the media address within the central network (2.2.2.2). The address returned by NAT device R (1.1.1.2) is then used in the OK message passed on to user
20 agent P.

User agent P then completes the SIP signalling sequence by sending an ACK message. This is passed along the chain of call agents to the called user agent V.

The user agents have each received an address within their local network to which media packets should be sent. This is the address of their local network address translation device, R and T. The NAT devices have been configured to send media packets received from the edge network 1 to the address of the other NAT device within
5 the middle network 2. The NAT devices have also been configured to send media packets received from the middle network to the address of the user agents in the edge network 1. Media packets can therefore be sent between the two user agents P,V via the two NAT devices R and T via media paths 3, 4, 5 in one direction, and via media paths 3a, 4a, 5a in the return direction.

10

The resulting media flow is looped unnecessarily through the central network 2, through paths 4, 4a.

It will be seen that IP calls traversing multiple NAT devices lose information about
15 preceding networks at each NAT device. If the call is routed back into a network segment that it has already traversed, then it is not possible to connect directly in that network segment. This can result in unnecessary network traffic and over-use of some network paths.

20 In one particular application, the edge network could be a small office, NATs T and R could be incorporated into one personal computer which can communicate with the internet 2, and user agents P and V could be other personal computers, and there could be further personal computers (not shown) in the edge network. The personal computers could all be in speech communication with each other using SIP, albeit via the personal

computer in communication with the internet. Unfortunately, each time a call is set up, such as between user agents P and V, it traverse NATs T and R , and uses up scarce media paths. The reason for going via the NAT pinholes is that the normal SIP protocol requires media addresses to be selected before the call agent determines the final destination of the call. If a route via the NAT is selected then the call will succeed if it terminates in a different network, and (by means of constructing another pinhole to re-enter the original network) will also succeed if the call terminates back in the same network as its origin. Selecting a NAT route for all calls is thus the safe option that will always work, albeit inefficiently.

10

A further example is shown with reference to Figure 2, which shows the set-up of a call using SIP signalling, between three networks. A call originating at device A in network 6 (Office LAN 1) is initially routed via the internet 7 to device B in network 8 (Office LAN 2), and then forwarded to device C in network 6. As the call traverses each NAT device 9,10 connecting the networks, information about the previous network is lost. The resulting call therefore consumes resources in all three networks. It will be noted that in Figure 2, both outward and return paths of the call are shown as a single line. This convention also applies to the other figures.

20 A solution known as dropback re-routing has been proposed using techniques such as the '302: Moved temporarily' response defined in SIP RFC3261, under which a new call is established directly between A and C, but this means that the call is no longer able to be controlled by devices elsewhere in the network (such as device B).

The invention provides a call set-up system, for the set-up of calls in a plurality of packet-switched networks connected to each other by network address translation (NAT) devices using a plurality of call agents, comprising means to send messages to successive call agents, which messages include address information for media packets
5 within networks associated with the call agents, to define the media path of the call, at least some of the messages also including address information for media packets sent to preceding call agents involved in the set-up of the call.

The system allows the media path of the call to take advantage of possible short-cuts
10 because call set-up does not only pass, as hitherto, address information for media packets within the network associated with the call set-up device, but also address information within networks associated with previous call set-up devices as well. Thus, if a call was set up from a first network to a second, and then back to the first, the messages passed would enable the media path of the call to be local to the first network
15 only rather than, as hitherto, traversing addresses in the second network. The present invention allows the route of associated media flows to be optimised. The call signalling, which consumes little network resource, could retain its original path.

Call set-up systems in accordance with the invention will now be described in detail, by
20 way of example, with reference to the accompanying drawings, in which:

Figure 3 which shows the set-up of a call using modified SIP signalling according to the invention;

Figure 4 shows the rules applied in the modified SIP signalling;

Figure 5 shows the set-up of a call using the modified SIP signalling for a series of networks, in which the call traverses several networks but does not re-enter an earlier
5 network;

Figures 6 to 8 show the set-up of calls using the modified SIP signalling for a series of networks, in which the call does re-enter an earlier network; and

10 Figure 9 shows the set-up of calls using the modified SIP signalling for a series of networks, in which the call does re-enter an earlier network, but in which various of the network address translation devices and their controlling call agents are legacy devices using unmodified SIP signalling.

15 All SIP signalling messages are shown in Figure 3, but only extracts from the text of messages is shown below. The basic format of SIP messages can be seen from IETF call examples documents.

As with the example of known SIP signalling described with reference to Figure 1,
20 network 1 could be a private network, and network 2 could be the internet.

User agent P initiates a call by sending an SIP Invite message to its local call agent Q. This message contains a session description indicating the media characteristics and the address (1.1.1.1) at which user agent P wishes to receive media packets. For

convenience, only the last two segments of this address, and of other addresses, are shown in Figure 3.

The full text of a typical Invite message is as follows, following a standard in IETF

5 documents:

INVITE sip:xxx@uax.middle.com SIP/2.0

Via: SIP/2.0/UDP uap.edge.com:5060;branch=z9hG4bK74bf9

Max-Forwards: 70

10 From: UserP ;tag=9fxced76sl

To: UserX

Call-ID: 3848276298220188511@uap.edge.com

CSeq: 2 INVITE

Contact: <sip:ppp@uap.edge.com

15 Content-Type: application/sdp

Content-Length: 151

v=0

o=ppp

20 s=-

c=IN IP4 1.1.1.1

t=0 0

m=audio 49172 RTP/AVP 0

a=rtpmap:0 PCMU/8000

Call agent Q responds with a SIP '100: trying' message.

Call agent Q determines that the destination of the call is in the central network 2, and that this is reached via a NAT device R which it controls. It therefore opens a pinhole in
5 NAT device R to permit a media flow from the central network to the media address of user agent P (1.1.1.1). The address returned by NAT device R (2.2.2.1) is then used in an Invite message sent to call agent S in the central network.

```
INVITE sip:xxx@uax.middle.com SIP/2.0
10  Via: SIP/2.0/UDP caq.edge.com:5060;branch=298bnsdhj2ka
    Via: SIP/2.0/UDP uap.edge.com:5060;branch=z9hG4bK74bf9
    Max-Forwards: 69
    From: UserP ;tag=9fxced76sl
    To: UserX
15  Call-ID: 3848276298220188511@uap.edge.com
    CSeq: 2 INVITE
    Contact: <sip:ppp@uap.edge.com>
    Content-Type: multipart/mixed
    Boundary: "****part separator****"
20  Content-Type: application/sdp
    Content-Length: 151

v=0

o=ppp
```

```
s=-  
c=IN IP4 2.2.2.1  
t=0 0  
m=audio 5378 RTP/AVP 0  
5 a=rtpmap:0 PCMU/8000  
  
--****part separator****  
Content-Type: application/localswitchstack  
NetworkID: edge.com  
10  
  
v=0  
o=ppp  
s=-  
c=IN IP4 1.1.1.1  
15 t=0 0  
m=audio 49172 RTP/AVP 0  
a=rtpmap:0 PCMU/8000  
--****part separator****
```

- 20 The SIP message syntax requires a string to be defined to separate the various 'attachments' to the message.

As with a typical SIP Invite message, the message contains the address "c=IN IP4 2.2.2.1" at which the NAT device R receives messages for transmission to User Agent P.

However, in accordance with the invention, call agent Q also places the previous session description within a stack structure as a multipart attachment to the SIP message, and this contains the network ID (in this case, "edge.com"), and the address in the edge network 1, "c=IN IP4 1.1.1.1", at which the User Agent P wished to receive
5 media packets. This is shown in Figure 3 in the first arrow in a descending direction extending from call agent Q to call agent S. The protocol relies on each address region (between NAT devices) having a globally unique identifier that can be recognised by all call set-up devices within that region. For SIP, this may in many cases derived from the domain name of the SIP server (as used in the SIP global call reference identifier).

10

Call agent S responds with a '100: trying' message.

Call agent S determines that user X has requested that their calls be forwarded to user agent V, and that the new destination is reached via call agent U in private network 1.

15 User X could be a home telephone number accessible via the internet, and the user could have set up an arrangement for their calls to be forwarded to their office in the private network 1. It therefore passes on the Invite with a changed URI. The stack structure described above is however retained.

20 Call agent U responds with a '100: trying' message.

Call agent U recognises that the call has arrived from a different network via NAT device T that it controls. It notices that the SIP message contains a local switching stack, and examines the Network ID values of the entries in the stack to see if the call has

passed through this network before. In this case it finds an entry for its own network ("edge.com"), and therefore uses the session description from this stack entry in the message it passes. No pinhole therefore needs to be opened in NAT device T. The Invite message sent to User Agent V just therefore contains the address (1.1.1.1) in network 1,
5 that is, just the local address, rather than the stack. This Invite message is as follows.

```
INVITE sip:vvv@uav.edge.com SIP/2.0
Via: SIP/2.0/UDP cau.edge.com:5060;branch=7tfvjyufh67
Via: SIP/2.0/UDP cas.middle.com:5060;branch=345hg2kuffs
10 Via: SIP/2.0/UDP caq.edge.com:5060;branch=298bnsdhj2ka
Via: SIP/2.0/UDP uap.edge.com:5060;branch=z9hG4bK74bf9
Max-Forwards: 67
From: UserP ;tag=9fxced76sl
To: UserX
15 Call-ID: 3848276298220188511@uap.edge.com
CSeq: 2 INVITE
Contact: <sip:ppp@uap.edge.com
Content-Type: application/sdp
Content-Length: 151
20
v=0
o=ppp
s=-
c=IN IP4 1.1.1.1
```

t=0 0

m=audio 49172 RTP/AVP 0

a=rtpmap:0 PCMU/8000

- 5 The call has now reached its destination. User agent V responds with a '180: ringing' message that is passed back to the caller via the chain of call agents.

When the call is answered, user agent V sends a '200: OK' message back to call agent U. The message contains a session description indicating the media characteristics and
10 the address (1.4) at which user agent V wishes to receive media packets. A typical message would read as follows.

SIP/2.0 200 OK

Via: SIP/2.0/UDP cau.edge.com:5060;branch=7tfvjyufh67

15 Via: SIP/2.0/UDP cas.middle.com:5060;branch=345hg2kuffs

Via: SIP/2.0/UDP caq.edge.com:5060;branch=298bnsdhj2ka

Via: SIP/2.0/UDP uap.edge.com:5060;branch=z9hG4bK74bf9

From: UserP ;tag=9fxced76sl

To: UserX

20 Call-ID: 3848276298220188511@uap.edge.com

CSeq: 2 INVITE

Contact: <sip:vvv@uav.edge.com

Content-Type: application/sdp

Content-Length: 151

v=0
o=vvv
s=-
5 c=IN IP4 1.1.1.4
t=0 0
m=audio 45678 RTP/AVP 0
a=rtpmap:0 PCMU/8000

- 10 Call agent U recognises that this message is for a call for which it invoked the local switching function. To allow compatibility with call agents that have not implemented the local switching functions (i.e. using unmodified SIP signalling), it constructs a stack containing a copy of the session description provided by user agent V, and sends this 200 OK message to call agent S, which passes it back to call agent Q.

15

- Call agent Q recognises that this call passes through NAT device R that it controls, and that the message contains a local switching stack. It therefore examines the stack to find entries for its own network. A matching entry is found, and is popped from the stack, that is, all higher entries on the stack are removed and discarded, the matching entry is
20 used as the new SDP, and any remaining lower entries are left on the stack, to form the session description in the SIP message that will be passed to user agent P. Since local switching has been invoked, the pinhole that was created during processing of the Invite message is no longer required, and is therefore deleted.

User agent P then completes the SIP signalling sequence by sending an ACK message. This is passed along the chain of call agents to the called user agent.

The user agents have each received an address within their local network to which
5 media packets should be sent. Since local switching has been invoked, this is the address of the other user agent. Single local media path 11 is used for the traffic. All superfluous pinholes within the NAT devices have been closed.

In normal SIP operation, the NAT controller re-writes the SDP (Session Description
10 Protocol (RFC 2616)) address information and discards the incoming address. If the call subsequently loops back into the originating network, the address information from that network has been lost and a local connection cannot be made. The scheme according to the invention with local switching avoids loss of information by pushing the previous session description onto a stack carried forward in the SIP signalling messages. When a
15 call enters a new network, the stack is scanned for entries previously inserted by this network. If an entry is found and local switching is permitted, then the stack is popped back to the original state in this network region. This means that the pinhole most recently allocated during re-entry to the region will not be required, and can be closed immediately.

20

It will be appreciated that the invention is an addition to SIP that permits flexible control of local switching where calls traverse a number of NAT devices. It should be noted that the SIP signalling performed by the user agents is unchanged by the presence of the local switching function, but that the media paths have been optimised. The

protocol is implemented by the call agents that control the transit of SIP media flows through the NAT devices, and these may be separate devices as in the embodiment of Figure 3, or incorporated in the NAT devices, as in the embodiments of Figures 5 to 9. Further, the call agents Q and U may be in the network 2 instead of in the network 1, if
5 desired. This could be useful in the case of a small office edge network, where call agents in the internet 2 could be used. The embodiments may also incorporate legacy call set-up devices such as in Figure 9, and such legacy devices need only transit the additional information generated by the call set-up devices.

10 While the description above has been in relation to SIP signalling, the principle may be extended to other signalling protocols based on offer/answer session descriptions to obtain the same benefits. Whatever the signalling protocol, the address region (between NAT devices) must have a globally unique identifier that can be recognised by all call set-up devices within that region.

15

For SIP, the identifier may in many cases derived from the domain name of the SIP server (as used in the SIP global call reference identifier). SIP uses an offer/answer protocol (RFC 3264) to convey bearer information. A session description is sent in each direction, indicating the media channels, addresses and codecs to be used. A NAT
20 control device (such as an ALG (Application Level Gateway)) re-writes the address information in the session descriptions to match the address translations configured in its local NAT device.

If features are present that require local switching to be inhibited they should simply delete the stack as they transit the message. This will force the media flow through the local network address. This approach could be applied independently for each direction of session description, but in some cases this would result in different NAT devices
5 being used for each direction, and would result in bypassed pinholes remaining open. Processing of the stack is therefore modified for the backward session description.

The invention will now be further explained with reference to Figures 4 to 9.

10 Figure 4 shows the rules and actions to be applied at each at each call agent. The rules should be checked in the order shown until a matching rule is found. It can be seen that the rules are the same in both directions, except that rules 1 and 3 will never apply for an 'offer' message, and need not be checked. The rule number used when processing a message at each NAT device is shown within a hexagon in each of the following
15 embodiments described with reference to Figures 5 to 9.

In the embodiments of Figures 5 to 9, the call agents are incorporated in the NAT devices, and are not shown separately. In the embodiment of Figure 5, there are five networks 12-16, joined by four NAT devices 17-20. The letters A and K represent
20 addresses of user terminals 21,22, and the letters B-J represent addresses of pinholes through the NAT devices. None of the networks is common to more than one pair of NAT devices. As with the embodiment of Figure 3, the call agents in the NAT devices place the previous session description within a stack structure as a multipart attachment

to the SIP Invite message sent to each successive NAT device. For example, the Invite SIP message sent to NAT device 18 includes the address C of the pinhole opened through the NAT device 17, but also the address A of the user terminal 21 in network 12 (see second arrow to the right in first line of arrows in Figure 5).

5

While the previous session description is passed along, there are no shortcuts possible for the media path between the terminals 21 and 22 because of the lack of common networks. Only Rule 4 applies. The address received in the session description is sent to the NAT device, and replaced by the address from the NAT device in the message
10 passed on by the call agent. This can be viewed as 'translating' the address within the SDP, and it must match the address translation that will be performed on media packets by the NAT device itself.

The embodiment of Figure 5 is to demonstrate that the invention does not disrupt calls
15 that traverse several networks but do not re-enter an earlier network.

Referring to the embodiment of Figure 6, there are three networks 23-25, joined by NAT devices 17-20. Network 23 is common to NAT devices 17 and 20, and network 24 is common to NAT device pairs 17,18 and 19,20.

20

Thus, when the SIP Invite message is passed from the call agent for network 25 to the call agent for network 24, Rule 2 applies, the stack of session descriptions is scanned through and replaced with the part of the stack that is headed by the earliest preceding

session description for the network 24. Thus, the stack of (G) [E,C,A] is replaced by (C)[A]. No pinhole is required to re-enter network 24, as indicated by circle 26.

Equally, the session description reverts to that of network 23 when the Invite SIP
5 message is passed to the call set-up device associated with re-entry into network 23. Rule 2 applies, and the stack of addresses (J)[C,A] is replaced just by (A) [], and the pinhole for re-entry into region 23 is closed. The media path of the call thus set up takes place wholly within the network 23, as path 27.

10 A pinhole is closed due to the operation of Rule 3 on the O.K. answer message. There are multiple Ks, not just one, in the answer path, arising from processing rules 1 & 3, which duplicate the session descriptions on the stack. It is needed in order to handle networks (like Figure 9) containing old call agent/NATs that do not implement the invention. Without this action the SDP could be lost when it passes through an old call
15 agent. However, when it is not lost this action results in multiple copies of the SDP on the stack. These are discarded during subsequent processing if the call loops back into a previous network.

In the embodiment of Figure 7, terminals and NAT devices 28-33 are joined by
20 networks 34 -36, and the resulting media path has three sections 37-39. The offer /answer messages, and the Rules applicable when processing a message at the NAT device, are shown in Figure 7.

The embodiment of Figure 8 has one further NAT device, so that the terminals and NAT devices are designated 40-46, and there are still only three networks 47-49, since the network 47 is common to both terminals 40,46, allowing a single local media path 50 for the call eventually set up. The offer /answer messages, and the Rules applicable
5 when processing a message at the NAT device, are shown in Figure 8.

The session description stack may reveal details of a call to other networks that some operators wish to keep private. Since the only requirement on the format of stack entries is that a network can traverse the stack and recognise its own entry, there is nothing to
10 prevent network devices from using encryption to ensure privacy of their stack data. Alternatively, networks may send a reference (URL) to the stack data as it appeared in their network in place of the session description. This requires that they hold the call state in a way that can be interrogated by any future call receiving the reference in the bearer info stack (e.g XML (Extensible Markup Language) document retrievable via
15 HTTP (Hypertext Transfer Protocol (RFC 2616))). The use of URLs will, however result in additional signalling between call handling devices in any network using this technique.

NAT controllers (call agents such as Q, T but not S) and SIP servers are already widely
20 deployed without support for the local switching technique described here. Figure 9 illustrates the impact that such devices may have on local switching. The embodiment is identical to that of Figure 8, but successive pairs of lines show the effect when various of the NAT devices are legacy equipment using conventional, not modified, SIP signalling. Thus, the first pair of lines shows the situation when the NAT device 41 is

legacy equipment. The Invite message to the call set-up device associated with network 48 simply replaces the address (A) with the address (C). The various applicable Rules when processing a message at the NAT device, are shown.

- 5 SIP devices (other than NAT controllers) may either transit the stack information unchanged, or delete the stack information. If the stack is transited unchanged, then local switching will operate normally. If the stack is deleted, then the media will be forced to traverse the network containing the incompatible device. Incompatible NAT controllers may transit the stack unchanged, or delete the stack. It is assumed that no
- 10 NAT controller would try to translate possible address fields in sections of the SIP message that it did not understand. If this did occur, other networks could prevent the problem by using simple data scrambling within stack entries. This would operate in the same way as the encryption technique described in the section on data hiding. If the stack is deleted, then the media will be forced to traverse the NAT controlled by the
- 15 incompatible controller. If the stack is transited unchanged, then local switching may be inhibited in some regions, and pinholes may remain open even if they are not subsequently used.

In all cases in Figure 9 both directions of media flow are set up, but the optimum local

20 switching path is not always found, and some unused pinholes are not closed.